



**Corela**

Cognition, représentation, langage

**HS-15 | 2014**

**Complémentarité des approches qualitatives et quantitatives dans l'analyse des discours**

---

## De la statistique textuelle à l'analyse des idéologies : l'exemple du discours révolutionnaire en Amérique latine (1810-2010)

Serge de Sousa

---



### Édition électronique

URL : <http://journals.openedition.org/corela/3554>

DOI : 10.4000/corela.3554

ISSN : 1638-573X

### Éditeur

Cercle linguistique du Centre et de l'Ouest - CerLICO

### Référence électronique

Serge de Sousa, « De la statistique textuelle à l'analyse des idéologies : l'exemple du discours révolutionnaire en Amérique latine (1810-2010) », *Corela* [En ligne], HS-15 | 2014, mis en ligne le 15 octobre 2014, consulté le 01 mai 2019. URL : <http://journals.openedition.org/corela/3554> ; DOI : 10.4000/corela.3554

---

Ce document a été généré automatiquement le 1 mai 2019.



Corela – cognition, représentation, langage est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International.

---

# De la statistique textuelle à l'analyse des idéologies : l'exemple du discours révolutionnaire en Amérique latine (1810-2010)

Serge de Sousa

---

## Introduction

- 1 L'articulation du qualitatif et du quantitatif est l'objet d'un questionnement déjà relativement ancien et qui a fait couler beaucoup d'encre hors des sciences du langage. C'est une question qui peut être posée, sous des modalités différentes, dans l'ensemble des sciences de l'homme. Même au sein d'une même « discipline<sup>1</sup> », la mise en œuvre des méthodes quantitatives ou des « méthodes mixtes » prend un nombre incalculable de formes. Dans le cadre de l'analyse du discours, particulièrement, le quantitatif et le qualitatif ont intérêt à être mobilisés ensemble en articulant « théorie du discours » (Pêcheux, 1969) et méthodes statistiques. C'est cette position que nous souhaiterions défendre à partir d'exemples d'analyses tirées de notre parcours de recherche. Pour ce faire, nous commencerons par situer notre approche sur le plan épistémologique en revenant brièvement sur la quantification des faits langagiers en histoire. Ensuite, c'est par l'exemple que nous illustrerons cette manière de faire, à partir de l'étude du discours révolutionnaire latino-américain.

# 1. Penser l'articulation du qualitatif et du quantitatif

## 1.1. Bref retour sur la progressive quantification des faits langagiers en histoire

- 2 Déjà anciennes en linguistique – Pierre Guiraud (1960 : 15) en faisait la « science statistique type » – les méthodes quantitatives sont mobilisées par diverses tendances – traitement automatique du langage, linguistique de corpus (Habert *et al.*, 1997), didactique (Dörnyei, 2007), sémantique (Rastier, 2011), lexicométrie (Lebart & Salem, 1994)... – parmi lesquelles l'analyse du discours. Celle-ci a ceci de particulier qu'elle a émergé au moment où le structuralisme d'une part et le quantitatif de l'autre étaient à l'apogée de leur influence au sein des sciences de l'homme ; la constitution de l'analyse du discours comme champ autonome au sein de la linguistique est directement issue, en France, des méthodes quantitatives – même si on peut distinguer de nombreuses autres influences (Mazière, 2005).
- 3 Anciennes aussi sont les relations entre *histoire* et *linguistique*<sup>2</sup> mais, dès leur développement, les méthodes de la statistique textuelle et de l'analyse du discours ont renouvelé ces relations par l'intérêt qu'elles ont soulevé auprès de certains historiens (Prost, 1969 ; 1974 ; Robin, 1973 ; Geffroy *et al.*, 1975). Mais l'histoire n'avait pas attendu l'analyse du discours pour mettre en œuvre des méthodes quantitatives. Elle avait, en son sein, indépendamment des sciences du langage, suscité des débats sur l'articulation du quantitatif et du qualitatif<sup>3</sup> – Pierre Chaunu (1978 : 9) parle de « grand débat des années 60 » – et appliqué mesures et dénombrements aux données économiques et démographiques<sup>4</sup>. Puis, c'est l'histoire sociale qui s'est vue traversée par le débat, à tel point qu'on en vint à affirmer que « scientifiquement parlant, il n'est d'histoire sociale que quantitative<sup>5</sup> ». Dans ce cadre, on assignait en particulier à l'histoire sociale la tâche « d'aboutir à une histoire de la psychologie collective », de « pénétrer les mentalités propres aux divers groupes sociaux » (Soboul 1967 : 21). Ainsi, un historien comme Chaunu, promoteur de l'histoire quantitative ou, comme il se plaisait à la nommer, *histoire sérielle*, percevait déjà, bien avant les expérimentations d'Antoine Prost, l'intérêt des méthodes quantitatives pour l'histoire des mentalités quand il faisait remarquer, après une application fructueuse (Chaunu & Chaunu, 1962), que :
 

« le mental collectif peut être codé, partant quand on a mis la main sur les bonnes séries documentaires, donnant lieu à une psychologie collective régressive. L'histoire, science auxiliaire, donne, ainsi, aux sciences de l'homme du présent cette épaisseur dans le temps, qui, lorsqu'elle est véritablement intégrée, constitue le substitut le plus sûr d'une impossible expérimentation » (Chaunu 1978 : 23).
- 4 Enfin, dans le même temps où les idéologies et les formations discursives étaient questionnées par des penseurs tels Louis Althusser, Michel Foucault ou Michel Pêcheux, l'histoire des idées se rapprochait d'une histoire des mentalités par la prise en compte de « la pensée automatique des organes d'opinion, [des] réflexes conditionnés, [de] la circulation des mythes et des stéréotypes, [des] nouveaux supports des énoncés idéologiques » (Winock 1996 : 247). Pour Foucault :
 

« l'histoire des idées s'adresse à toute cette insidieuse pensée, à tout ce jeu de représentations qui court anonymement entre les hommes ; dans l'interstice des grands monuments discursifs, elle fait apparaître le sol friable sur lequel ils reposent. C'est la discipline des langages flottant, des œuvres informes, des thèmes

non liés. Analyse des opinions plus que du savoir, des erreurs plus que de la vérité, non des formes de pensées mais des types de mentalités » (1969 : 179).

- 5 À cette époque où l'analyse du discours prenait forme en associant modèle linguistique et modèle sociologique, cherchant à « comprendre la part de la variation des langages dans les pratiques sociales » à travers « la co-variance entre des structures linguistiques et des modèles sociaux » (Guilhaumou 2006 : 13), la lexicométrie, s'appuyant sur les travaux de la linguistique quantitative de Pierre Guiraud et Charles Muller, s'est alors imposée, autour de Maurice Tournier et du groupe de Saint Cloud, comme un moyen de quantifier des faits langagiers. Et quelques années plus tard, en dépit de la méconnaissance des historiens pour les analyses statistiques des données textuelles et les analyses langagières – bien que certains d'entre eux avaient pris part activement à l'émergence de l'analyse de discours et de la lexicométrie –, Prost mettait en garde ses confrères contre l'« impasse » que constitue pour l'historien la « multiplication de citations » qui, bien qu'elles aient « une valeur illustrative », « ne constitue jamais une preuve irréfutable » (Prost 1996 : 258-259). « À cette impasse – disait-il – un seul moyen d'échapper, semble-t-il : les méthodes quantitatives comme l'analyse de contenu ou la statistique lexicale. La fréquence des termes, des expressions, des opinions ou des jugements paraît un indicateur sûr de leur importance objective » (Prost 1996 : 258-259), avant d'ajouter peu après :

« Les façons de parler ne sont pas innocentes ; par delà leur apparente neutralité, elles révèlent des structures mentales, des façons de percevoir et d'organiser la réalité en la nommant. Elles trahissent les préjugés et les tabous par leurs stéréotypes ou leurs silences. Bref à côté des démarches ethnologiques, les démarches linguistiques constituent l'une des voies les plus sûres d'une histoire des mentalités.

Dans cette perspective, l'attention se porte de façon prioritaire sur les structures lexicales et sémantiques. La fécondité de la statistique lexicale tient d'ailleurs à ce qu'elle met en évidence les termes forts de ces structures, ceux qui leur servent de pivots, fondant ainsi sur des critères quantitatifs incontestables une étude plus qualitative. » (Prost 1996 : 270-271)

- 6 C'est dans cette voie, tracée par Prost, que l'on peut inscrire l'articulation du quantitatif et du qualitatif dans l'analyse de discours en histoire sociale, et particulièrement en ce qui concerne l'analyse des idéologies. L'histoire – et plus spécifiquement l'histoire du discours, de la pensée politique ou l'histoire culturelle – a donc beaucoup à gagner de l'utilisation de protocoles d'analyse quantitative des textes (Duchastel & Armony, 1993 ; Armony, 1997 ; 2005), en particulier en ce qui concerne la question, centrale en histoire, de « l'administration de la preuve » (hypothético-déductif), mais aussi pour suggérer des pistes de recherches (heuristique) qui, à l'œil nu, n'auraient peut-être pas attiré l'attention de l'historien (Prost, 1996 ; Mayaffre, 2002 ; 2005).

- 7 Mais la complémentarité du quantitatif et du qualitatif dans l'étude des discours et des idéologies ou dans l'articulation, comme le suggère Prost, de la statistique lexicale et de l'histoire, doit-elle encore, pour être féconde, s'appuyer sur un arrière-plan conceptuel fort que seule peut fournir une « théorie du discours » (Pêcheux, 1969) définie comme :

« une analyse linguistique, articulée à la fois aux genres discursifs, aux formations discursives, aux situations concrètes de communication et aux pôles énonciatifs [...] pren[ant] en charge des corpus larges et variés de discours authentiques, rapport [ant] ces derniers à leurs conditions sociohistoriques de production et de circulation, mett[ant] à jour leur idéologie sous-jacente » (Détrie *et al.*, 2001 : 7-8).

- 8 À ce stade, l'important tant pour l'analyse du discours et la lexicométrie que pour l'histoire quantitative, « c'est la constitution du matériau d'analyse » (Furet 1982 : 60). Nous ne reviendrons pas ici sur la question de la constitution des corpus dont les impératifs ont déjà été explicités par d'autres auteurs à qui nous renvoyons (Mayaffre 2000 : 38-42 ; Guilhaumou, 2002). Indiquons seulement que les procédures quantitatives requièrent un travail de réflexion sur les données à analyser qui amène, « oblige » dit Furet, l'historien à « renoncer à la naïveté méthodologique et [à] réfléchir aux conditions d'établissement de son savoir » (Furet 1982 : 64). En effet, dit autrement par l'analyste du discours, « c'est souvent le corpus qui, en fait, définit l'objet de recherche qui ne lui préexiste pas. Ou plutôt, c'est le point de vue qui construit un corpus, qui n'est pas un ensemble prêt à être enregistré » (Charaudeau & Maingueneau (dir.) 2002 : 149). Une fois de plus disparaît l'illusion qu'une analyse quantitative du discours peut s'extraire de toute considération qualitative.

## 1.2. Mesurer et interpréter

- 9 Aussi nous postulerons – et nous illustrerons – une complémentarité de fait entre le quantitatif et le qualitatif en analyse du discours, dont l'une des modalités peut être, nous venons de le dire, l'articulation d'une théorie du discours et de ses concepts à un protocole d'analyse statistique des données textuelles au sein d'une histoire sociale ou culturelle. Plus généralement, dès le moment où il y a intervention du chercheur autre que pour appuyer sur un simple bouton, toute approche dite « quantitative » combine de fait quantitatif et qualitatif. Dès que le chercheur pénètre les résultats quantitatifs obtenus, qu'il les ordonne, les assimile, et à plus forte raison, qu'il en rend compte ou les interprète, dès, en fin de compte, qu'il y a « lecture », il y a appréciation « qualitative » des résultats quantitatifs (Fiala, 1994). Qualitatif et quantitatif sont donc le plus souvent articulés – parfois même inconsciemment, c'est-à-dire sans parti-pris méthodologique. Bien entendu, on peut considérer que certaines recherches privilégient le quantitatif quand d'autres s'efforcent de l'ignorer ou de l'évacuer et que, pour des raisons méthodologiques ou les besoins de la réflexion épistémologique, il est possible de les distinguer ou encore de les doser. Cependant, il ne faut pas tomber dans l'illusion d'une totale indépendance de ces deux paradigmes, pour le moins en analyse du discours.
- 10 L'interprétation constitue d'ailleurs le moment le plus délicat de la démarche lexicométrique car elle « transforme des constats en opinions » (Tournier 1980 : 200). Il s'agit, après avoir effectué toute une série de mesures et dénombrements, de mettre en rapport les constats statistiques et quantitatifs avec les hypothèses sociopolitiques ayant motivées la constitution du corpus. Interpréter en lexicométrie, c'est valider ces hypothèses sociopolitiques « par une suite d'inférences remontant des constats statistiques aux usages discursifs et linguistiques dont ils sont la trace, puis aux données socio-historiques qui en constituent les causes » (Fiala 1994 : 119). On réintroduit donc la subjectivité du chercheur et l'appréciation qualitative là où l'on s'était efforcé de la neutraliser en constituant une démarche objectivable et reproductible. Là repose donc, principalement, l'articulation du quantitatif au qualitatif en lexicométrie.
- 11 Mais entre la mesure du corpus et l'interprétation sociohistorique, la théorie du discours permet de redonner toute son épaisseur à la formation discursive étudiée à travers l'utilisation de concepts qui, associés à certaines méthodes quantitatives et statistiques, apparaissent particulièrement féconds pour témoigner de la complémentarité du

quantitatif et du qualitatif. Ainsi en est-il, par exemple, des notions telles que « formation discursive », « stéréotype », « formule » ou « interdiscours » qui, articulées<sup>6</sup> aux méthodes probabilistes et quantitatives de la statistique textuelle – dans le cas présent le calcul des cooccurrences, les proximités segmentales et les analyses multidimensionnelles –, permettront d'étudier quelques exemples du discours révolutionnaire latino-américain.

- 12 Pour illustrer cette complémentarité entre quantitatif et qualitatif, nous montrerons, à partir d'un corpus regroupant les principaux textes et manifestes des grandes figures révolutionnaires d'Amérique latine dont la renommée a largement dépassé les frontières de leur propre pays, soit 452 textes produits par 23 locuteurs latino-américains entre 1811 et 2009, une « manière de faire en analyse du discours » à travers trois exemples : l'analyse de l'un des principaux stéréotypes cimentant la doxa révolutionnaire, la mise au jour de l'interdiscours marxiste et la constitution d'une généalogie entre les différents locuteurs révolutionnaires qui constituent le corpus.

## 2. Mettre en œuvre la complémentarité du qualitatif et du quantitatif

- 13 Le tableau ci-dessous présente les principales caractéristiques du corpus d'analyse partitionné par locuteurs. On y trouve des leaders indépendantistes, des dirigeants des Partis socialiste ou communiste, ainsi que différentes guérillas. Lors de certaines expérimentations (voir *infra* §2.2), on ajoutera à ce corpus les principaux textes et ouvrages des sept principales figures tutélaires du marxisme européen et asiatique.

Locuteurs	Période	Textes	Occurrences	Type	Pays
1 Bolivar	1811-1830	11	36556	Révolution d'indépendance	Venezuela
2 Martí	1880-1895	21	57111	Révolution d'indépendance	Cuba
3 Recabarren	1904-1925	17	47839	Fondateur <i>Partido Comunista</i>	Chili
4 Zapata	1911-1919	42	50302	Révolution nationale	Mexique
5 Mella	1923-1928	20	39651	Fondateur <i>Partido Comunista</i>	Cuba
6 Mariategui	1924-1929	13	79223	Fondateur <i>Partido Socialista</i>	Pérou
7 Sandino	1927-1933	29	41566	Révolution nationale	Nicaragua
8 M26	1953-1957	7	39795	Guérilla	Cuba
9 Guevara	1959-1967	13	73209	Guérilla, homme d'État	Cuba
10 Castro	1960-1983	4	47375	Guérilla, chef d'État	Cuba
11 Torres	1956-1966	26	44233	Guérilla, théologie de la libération	Colombie
12 FSLN	1960-1978	40	83038	Guérilla	Nicaragua
13 MLN-Tup.	1967-1987	11	38606	Guérilla (urbaine)	Uruguay
14 Allende	1970-1973	10	64933	Fondateur <i>Part. Socialista</i> , chef d'État	Chili
15 Santucho (ERP)	1968-1976	14	52785	Guérilla	Argentine
16 FMLN	1980-1990	8	53589	Guérilla	El Salvador
17 MRTA	1982-2007	13	46342	Guérilla	Pérou
18 PCP-SL	1988-2008	20	61426	Guérilla	Pérou
19 EZLN	1994-2006	10	60547	Guérilla	Mexique
20 EPR	1994-2001	16	63556	Guérilla	Mexique
21 FARC	1994-2009	29	61600	Guérilla	Colombie
22 ELN	1998-2008	35	48099	Guérilla	Colombie
23 Chavez	2009	43	50522	Chef d'État	Venezuela

Locuteurs	Période	Textes	Occurrences	Locuteurs	Période	Textes	Occurrences
24 Engels	1847-1894	5	46885	28 Gramsci	1917-1931	15	39466
25 Marx	1848-1872	3	52707	29 Trotski	1930-1938	2	71356
26 Lénine	1899-1906	3	66476	30 Mao	1938-1964	4	61629
27 Staline	1913-1924	2	55066				

Corpus (1811-2009)	Locuteurs	Textes	Taille
RévolutionAmLat	23	452	1241903
Révolution	30	486	1635488

Tableau 1 : Principales caractéristiques du corpus d'étude.

## 2.1. Cooccurrences et stéréotypes : les États-Unis dans le discours révolutionnaire latino-américain

- 14 À travers la consultation du premier document lexicométrique, l'*index* (hiérarchique ou alphabétique), un paradigme apparaît comme important dans ce discours : les *États-Unis* (« *estados unidos* »). On constate que ce paradigme s'actualise dans le corpus à travers trois sous-paradigmes *ESTADOS UNIDOS*, *NORTEAMÉRICA* et *YANKI*, chacun de ceux-ci étant ensuite déclinés sous ses formes nominales et adjectivales<sup>7</sup> (*Estados Unidos*, *estadounidense*, *norteamericanas*, *yankis*, etc.). Le paradigme, à travers les 13 items recensés (figure 1), est attesté à 1825 reprises<sup>8</sup> dans le corpus et, au vu de la fréquence d'autres acteurs ou concepts sociopolitiques dans le corpus<sup>9</sup>, et en particulier du paradigme de l'Amérique latine, les États-Unis se situent en bonne place dans l'ordre des préoccupations des révolutionnaires.

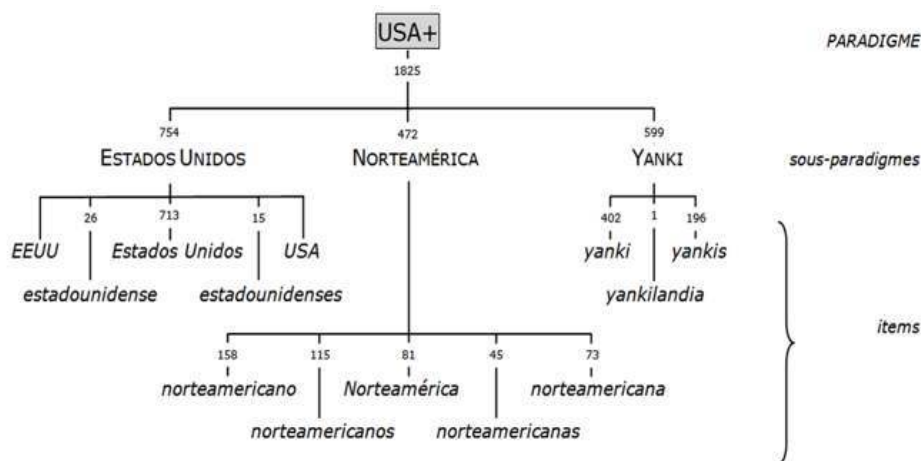


Figure 1 : Paradigme des États-Unis<sup>10</sup> dans le corpus.

- 15 Ici, la première des trois méthodes dont l'intérêt pour l'historien est certain, le calcul des cooccurrences<sup>11</sup> – c'est-à-dire de l'« ensemble des relations d'attraction, d'association ou de répulsion qu'un vocable entretient avec les autres composant le lexique d'un corpus » (Labbé & Monière 2003 : 26), calculées par le modèle probabiliste –, se révèle un atout précieux. Avec celle-ci, l'analyse quantitative fait un pas en avant – voire un saut – vers l'analyse qualitative. En mettant en lumière les réseaux sémantiques tissés autour d'une forme-pôle, cette méthode permet d'approcher plus aisément le sens du mot en discours ; même si la consultation des contextes demeure nécessaire, elle s'avère cependant plus aisée (on ne consulte plus les contextes uniquement sur la base d'une forme mais à partir de la présence conjointe de deux formes, voire, avec le logiciel Coocs de William Martinez, de trois, quatre, cinq formes...).
- 16 Surtout, le passage de la statistique textuelle à l'analyse de discours, puis à l'histoire des mentalités ou à la psychologie sociale peut se faire en passant des tableaux et graphiques de cooccurrences à l'analyse d'éventuelles cristallisations de la pensée amenant à la détection de stéréotypes<sup>12</sup>, propres à un positionnement politique ou à un groupe social. L'analyse qualitative des résultats cooccurentiels, quantitatifs, peut faciliter *in fine* l'analyse des représentations sociales et des mentalités (De Sousa, 2012a). Et les résultats, dans le cas du corpus révolutionnaire étudié ici, sont particulièrement fructueux : on calcule les associations positives, statistiquement significatives, entre notre paradigme et



les autres formes du corpus dans le discours révolutionnaire, puis, par commodité, ces associations sont représentées par un graphique :

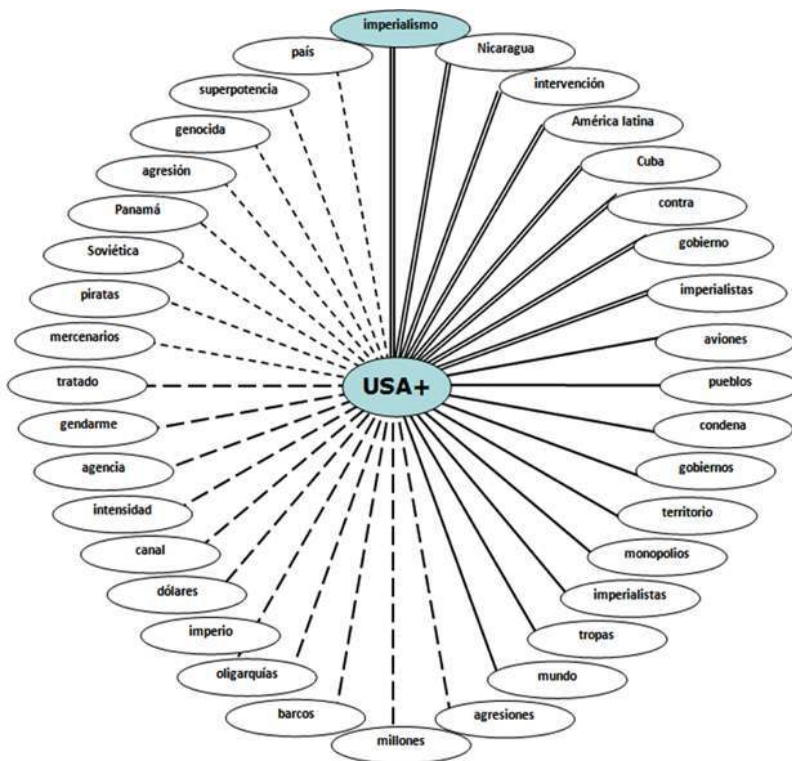


Figure 2 : Univers lexical des États-Unis<sup>13</sup>

- 17 La figure est éminemment significative. En effet, la première constatation – statistique, rappelons-le – vaut tout un discours : le principal co-occurent – le mot le plus fréquemment associé au paradigme des États-Unis dans le corpus révolutionnaire – est *imperialismo*. La suite des co-occurents n'est que le déroulement de cette même image – « les États-Unis, c'est l'impérialisme » – et cette longue liste ne cesse de surprendre par son homogénéité : *intervención, gobierno, imperialistas, condena, monopolios, tropas, agresiones, imperio, mercenarios, agresión...*
- 18 On distingue en effet, en filigrane, à travers la liste des co-occurents la définition de l'impérialisme : une puissance interventionniste (co-occurents : *intervención, agresión, tropas...*) et monopolistique (*dólares, monopolios, millones...*). Ces deux traits, qui constituent la conception léniniste de l'impérialisme comme « stade suprême du capitalisme », construisent l'image des États-Unis qui, par la récurrence des co-occurents<sup>14</sup> se fige, dans le discours révolutionnaire latino-américain, en stéréotype<sup>15</sup>, exprimant un imaginaire social propre.
- 19 C'est l'impérialisme politique, et militaire, qui, de loin, est le plus présent dans le corpus, construit à partir de l'évocation récurrente des *interventions* étasuniennes en *Amérique latine* (*Nicaragua, Cuba, Panamá...*), même si la « diplomatie du dollar », dénoncée comme « la explotación imperialista norteamericana a los pueblos subdesarrollados de la América latina », complète le portrait fortement négatif des États-Unis. En explorant les contextes, on retrouve ces « représentations toutes faites, [ces] schèmes culturels préexistants, à l'aide desquels chacun filtre la réalité ambiante », qui « expriment un imaginaire social » (Amossy & Herschberg Pierrot 2009 : 26) et médiatise le rapport au réel, en particulier à



partir de la circulation de formules comme « impérialisme yanki » ou « intervention yanki » dont la présence apparaît rapidement à la lecture des contextes :

20 - Interventionnisme politique et militaire :

« Panamá, Puerto Rico, Cuba, Haití, Santo Domingo y Nicaragua, son las seis desgraciadas repúblicas que *perdieron su independencia* y que han pasado a ser colonia del **imperialismo yanki**. » (Sandino)

« La Asamblea General Nacional del Pueblo de Cuba, condena enérgicamente la **intervención** abierta y criminal que durante más de un siglo ha ejercido el **imperialismo norteamericano** sobre todos los pueblos de la América Latina, pueblos que más de una vez han visto invadido su suelo en México, Nicaragua, Haití, Santo Domingo o Cuba, que han perdido ante la voracidad de los **imperialistas yankis**, extensas y ricas zonas como Tejas, centros estratégicos vitales como el canal de Panamá, países enteros como Puerto Rico convertido en territorio de ocupación ; que han sufrido, además, el trato vejaminoso de los infantes de marina, lo mismo contra nuestras mujeres e hijas que contra los símbolos más altos de la historia patria como la efigie de José Martí. » (Castro)

« Las **agresiones norteamericanas** no solamente causan destrozos en lo material, sino que también repercuten en el patrimonio cultural y espiritual de los pueblos. Solamente así puede explicarse el olvido en que han caído muchas de las **intervenciones de los Estados Unidos** en Nicaragua [...] » (FSLN)

« Para ello un primer paso es exigir el **retiro de las bases militares y de cualquier forma de presencia militar norteamericana de nuestro país y de América latina**, de tal forma que los pueblos ejerzan su soberanía sin amenazas de ninguna índole » (MRTA)

« [...] pida y obtenga, aunque sea a balazos, el **retiro inmediato de las fuerzas militares de ocupación norteamericanas en Nicaragua** » (Sandino)

« [...] todo obedece a consignas de la **política expansionista de los Estados Unidos de Norteamérica** » (Sandino)

21 Les calculs statistiques et quantitatifs ont fait aussi ressortir, bien qu'au second plan, l'aspect économique de l'impérialisme : le monopole (*monopolios, dólares, millones, oligarquías...*) :

22 - Dénonciation du capitalisme et des monopoles :

« La implantación por los Estados Unidos, de manera irresponsable e inconsulta, de altas tasas de interés, en la búsqueda de egoístas objetivos económicos nacionales, repercutió directamente en el Tercer mundo, cuyo servicio de la deuda externa ascendió, a finales de 1982, a la impresionante cifra de unos 131000 millones de dólares.

En lo que se refiere a las **transnacionales norteamericanas**, entre 1970 y 1979 invirtieron 11446 millones de dólares y **repatriaron en forma de ganancia 48663 millones, lo que significa 4,25 dólares extraídos del Tercer mundo por cada nuevo dólar invertido** en ese período. » (Castro)

23 De plus, l'analyse et l'histoire du discours tirent encore profit des procédés lexicométriques quand ceux-ci leur permettent d'identifier et de suivre les reformulations et la circulation d'une formule-clé du stéréotype des États-Unis dans le discours révolutionnaire latino-américain : *intervención (directa)* devient *intromisión en nuestros asuntos internos* ou encore *incesante hostigamiento de intervención yanki en nuestros asuntos internos*, *intervención del imperialismo yanki*, la **intervención directa del imperialismo yanqui**, la *intervención abierta y criminal que durante más de un siglo ha ejercido el imperialismo norteamericano*, circulant entre les discours de Sandino, Castro ou du Front Sandiniste de Carlos Fonseca :

« Hasta entonces los inmorales **gobiernos** de los **Estados Unidos de Norteamérica** habían conseguido mantener en Nicaragua su política cubierta con una máscara de

hipocresía, que les permitía presentarla ante el mundo, al simple golpe de vista, con un aspecto de legalidad, *encubriendo así su intromisión en nuestros asuntos internos.* » (Sandino)

« Cuba ha vivido tres años de revolución bajo *incesante hostigamiento de intervención yanqui en nuestros asuntos internos.* » (Castro)

« Por su parte el **imperio norteamericano** invadía los países a su antojo e imponía gobiernos militares, con la excusa de defender sus intereses. » (ELN)

- 24 Surtout, si le stéréotype véhiculé par le discours révolutionnaire se construit bien à partir de la répétition des mêmes vocables – mis au jour par le calcul des co-occurents (impérialisme, intervention, agression, monopoles...) –, on constate à partir des contextes consultés que le réseau isotopique dépasse largement les seuls co-occurents spécifiques relevés et se développe à travers toute une série de termes aux sèmes communs qui n'apparaissent pas immédiatement dans la liste des co-occurents : *conquista, perder su soberanía, perder su independencia, luchar, penetración económica y política, amenazar la soberanía territorial, invadir su suelo, voracidad...*
- 25 La diachronie étant l'essence de l'histoire, le graphique ci-dessous montre l'évolution de la présence au cours des ans du paradigme des États-Unis dans le corpus.

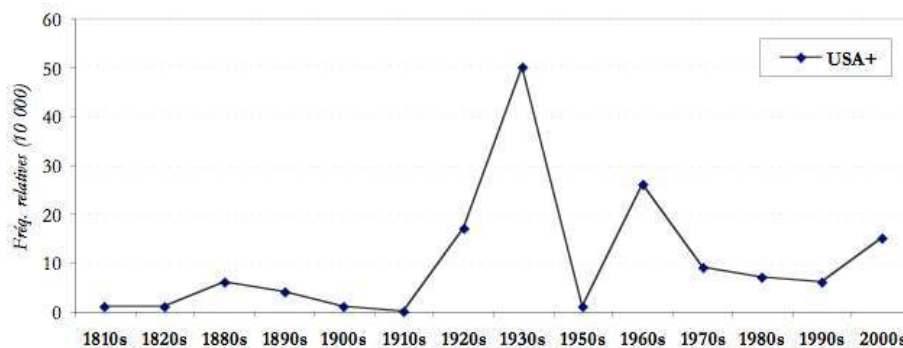


Figure 3 : Ventilation du paradigme (fréquences relatives x10 000)

- 26 On constate trois temps dans l'utilisation du paradigme. Une première phase – de 1810 à 1910 – qui connaît une faible utilisation du paradigme : c'est l'époque qui va de Bolivar à Zapata, en passant par Recabarren et Martí. Bien entendu cette période connaît en elle-même des fluctuations, notamment à travers les textes de Martí, plus marqués par la présence étatsunienne. Une deuxième phase débute en 1920 pour aller jusqu'en 1960 et se distingue par une forte utilisation du paradigme. Enfin, dans la troisième phase, à partir des années 70 jusqu'au 21<sup>ème</sup> siècle, le paradigme est utilisé dans des proportions moyennes.
- 27 Ces variations quantitatives recourent, pour sûr, des variations qualitatives dans la dénomination des États-Unis. Le même graphique prenant en compte cette fois non pas le seul paradigme mais les trois sous-paradigmes individuellement, apporte un éclairage supplémentaire sur l'évolution de l'image des États-Unis et permet de dater l'apparition du stéréotype :

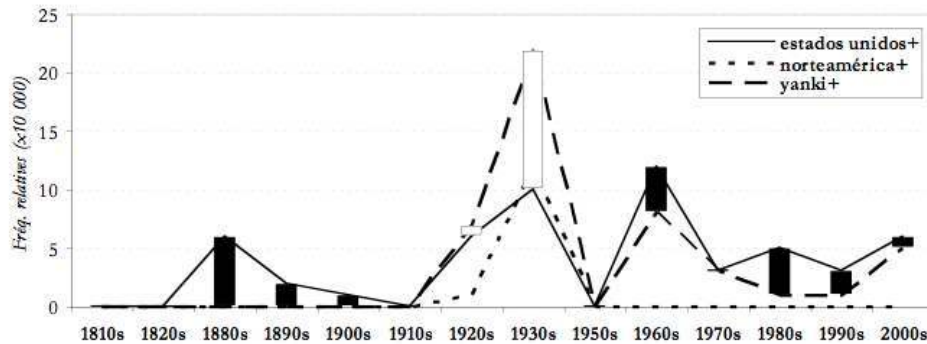


Figure 4 : Ventilation des trois sous-paradigmes (fréquences relatives x10 000).

- 28 Dans notre corpus, les États-Unis et ses habitants sont nommés à travers trois désignants principaux : tout d'abord *Estados Unidos+* (le plus fréquent et le plus neutre), *yanki+* (deuxième dans l'ordre des fréquences et le plus péjoratif), et *Norteamérica+* (nettement moins utilisé que les deux autres et dont le sème [nord] vs [sud] apparaît beaucoup plus marqué que dans les deux autres syntagmes et comme un élément non négligeable de son sémantisme). On constate grâce à la figure 4 que, concomitamment à l'évocation de plus en plus fréquente des États-Unis dans le discours révolutionnaire latino-américain, correspondant à l'apogée de l'impérialisme étatsunien, la stéréotypisation commence à s'installer à travers le désignant *yanki* (et dans une moindre mesure, *norteamérica*). À partir des années 20, dans le discours révolutionnaire latino-américain, les désignants *yanki* et *norteamérica* prennent le pas sur la dénomination neutre et officielle de *Estados Unidos* pour mettre en avant l'altérité de cette figure à travers la prégnance du sème [norte] et la péjoration à l'œuvre dans *yanki*. Le passage de *Estados Unidos* à *norteamérica* et *yanki* peut être vu comme le figement de la représentation des États-Unis dans le rôle du voisin du Nord à combattre. Les États-Unis, dans le discours révolutionnaire à partir de 1920, sont vus à travers le prisme du stéréotype du yankee.
- 29 Enfin, si l'on s'intéresse à la significativité statistique de ces variations de fréquence, à partir du modèle probabiliste, on peut dire que les années 30 et les années 60 sont les deux périodes qui surinvestissent fortement le thème des États-Unis, ce qui confirme que cette thématique prend pleinement part aux idéologies des combats.
- 30 Sans pousser plus loin l'analyse, on voit comment, à partir de cet exemple, le calcul statistique des co-occurents suggère que la représentation des États-Unis dans le discours révolutionnaire latino-américain s'apparente à une critique marxiste de l'impérialisme, articulant interventionnisme politique et monopolisme économique, et qu'elle se constitue en doxa<sup>16</sup>. Si l'on a pu déceler – à partir de la présence au premier plan des États-Unis et grâce au calcul des co-occurents – le substrat idéologique du discours révolutionnaire latino-américain, l'utilisation d'une deuxième méthode lexicométrique permettra de préciser cet arrière-plan idéologique en fonction des mouvements révolutionnaires.

## 2.2. Proximités segmentales et interdiscours marxiste

- 31 Le deuxième outil de la statistique textuelle d'un intérêt certain pour l'analyse de discours, l'analyse qualitative ou l'étude historique est l'*inventaire des segments répétés* : la statistique textuelle permet aisément de dresser la liste des segments phrastiques qui apparaissent à plusieurs reprises dans un corpus, puis de visualiser leur ventilation entre

les parties du corpus. Associés à des notions fécondes en analyse du discours comme celles, parmi d'autres, de *formule*<sup>17</sup> ou d'*interdiscours*<sup>18</sup>, la méthode des segments répétés permet d'étudier la circulation des formules tant diachroniquement qu'entre plusieurs locuteurs et, *in fine*, de détecter quantitativement l'interdiscours sur de grands corpus (De Sousa, 2010). Ici, la statistique textuelle peut grandement contribuer à l'histoire des idées ou aux recherches sur la circulation des savoirs : on l'a aperçu pour la formule « imperialismo yanqui », on peut le voir plus concrètement à partir des formules marxistes.

- 32 Reprenons notre corpus révolutionnaire. À partir du paradigme des États-Unis, on a pu voir qu'une critique de l'impérialisme, idéologiquement proche du marxisme-léninisme, prenait corps. La méthode des segments répétés peut permettre d'évaluer la présence de l'interdiscours marxiste dans le corpus à l'aide par exemple de la notion de *proximités segmentales*. La notion de proximités segmentales, développée par André Salem, naît d'une interrogation sur la circulation, à l'intérieur du discours politique, de lexies figées caractéristiques de certains corpus idéologiques :

« l'attestation simultanée dans deux textes de séquences identiques dont la taille dépasse la forme induit le sentiment que les textes se réfèrent à des concepts communs, qu'ils ont peut-être été produits dans des conditions proches, sinon par des *formations discursives* proches, même si leurs stocks lexicaux, pris dans leur totalité, ne présentent pas de similitudes particulières » (Salem 2006 : 841).

- 33 Introduisons dans notre corpus révolutionnaire sept nouveaux sous-corpus, contenant pour chacun les textes fondamentaux des figures tutélaires du marxisme. Le calcul des segments répétés permet, à partir de l'inventaire dressé par *Lexico 3*, de réunir dans un groupe de formes une série de formules et de désignants marxistes à fort contenu idéologique et particulièrement fréquents dans le corpus. Chacun des trente segments répétés retenus est attesté chez plusieurs locuteurs marxistes et au moins 50 fois dans l'ensemble du corpus.

Segments	F	Segments	F	Segments	F
la lucha revolucionaria	116	la lucha de clases	205	la lucha popular	67
la lucha económica	82	la guerra de guerrillas	220	la guerra popular	414
la guerra revolucionaria	73	la revolución proletaria	182	la revolución socialista	134
la revolución proletaria mundial	55	las fuerzas productivas	67	las fuerzas populares	64
las fuerzas revolucionarias	104	la clase trabajadora	57	la clase popular	102
las clases dominantes	62	la clase dirigente	77	la clase obrera	463
el movimiento revolucionario	62	el movimiento obrero	57	las masas populares	110
la dictadura del proletariado	188	la dictadura democrática	60	la dictadura militar	60
la democracia burguesa	56	la conquista del poder	125	la toma del poder	119
los medios de producción	87	la propiedad privada	77	la pequeña burguesía	127

Tableau 2 : Liste des segments répétés « formules marxistes »

- 34 Le graphique des spécificités confirme que ces formules sont représentatives du discours marxiste (spécificités  $\geq 28$  pour Staline, Trotski, Gramsci et Mao<sup>19</sup>) et permet de distinguer parmi les locuteurs latino-américains ceux qui les reprennent le plus fréquemment : *Santucho* et *Sentier Lumineux* d'une part, *Torres* et *EPR* d'autre part.
- 35 A contrario, le discours des Révolution d'Indépendance et des Révolutions nationales ( *Bolívar* [1], indépendantiste, antérieur à l'émergence du marxisme, mais aussi José Martí [2], Emiliano Zapata [4] ou Augusto C. Sandino [7]) ne se situent pas du tout sous l'influence idéologique du marxisme. C'est surtout la Révolution cubaine (Guevara [9], Castro [10]) qui marque une rupture qualitative avec les locuteurs antérieurs. Même s'ils sont loin d'être les plus fervents utilisateurs des formules marxistes – ils les sous-utilisent

par rapport à certains de leurs successeurs et surtout par rapport aux figures tutélaires du marxisme – ce sont eux, Guevara et Castro qui initient un changement dans le discours révolutionnaire latino-américain – Guevara plus que Castro d'ailleurs. Enfin, on distingue deux mouvements révolutionnaires dont la phraséologie apparaît statistiquement très marquée par l'idéologie marxiste : Mario Santucho (EGP, Argentine) et Sentier Lumineux (Pérou).

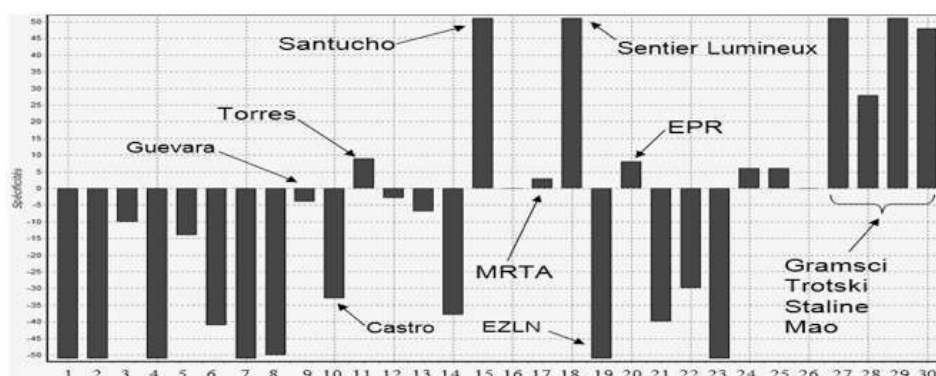


Figure 5 : Formules marxistes dans le discours révolutionnaire.

- 36 Comme on le voit, le travail sur les proximités segmentales et sur la circulation des formules offre des pistes d'investigation pour l'étude de l'influence et de la confluence entre les mouvements révolutionnaires d'une part, et ces mêmes mouvements et le marxisme d'autre part, en particulier dans la mise au jour du substrat idéologique des discours. L'étude de la confluence et des parentages<sup>20</sup> entre les mouvements révolutionnaires peut bénéficier d'un troisième type de méthodes de la statistique textuelle : les analyses multidimensionnelles.

### 2.3. Analyses multidimensionnelles et parentages idéologiques

- 37 Troisième type de méthodes quantitatives particulièrement fécondes pour l'analyse qualitative, les méthodes multidimensionnelles (AFC, arborée). Celles-ci permettent de rapprocher les parties d'un corpus en fonction de la quantité de vocabulaire<sup>21</sup> qu'elles partagent. Plus deux parties auront un vocabulaire proche, plus les classifications les rapprocheront permettant de dégager des typologies. À partir d'une série textuelle chronologique, qui peut couvrir quelques mois comme plusieurs dizaines, voire centaines d'années, il est possible, grâce aux outils statistiques, d'en dégager la structuration interne en fonction tant du vocabulaire que des catégories grammaticales, des enchaînements syntaxiques ou d'un groupe de formes ou de thèmes. L'historien est particulièrement concerné par la diachronie et dans ce cadre, les « séries » se révèlent un point de jonction entre histoire et statistique. Les résultats obtenus doivent être analysés minutieusement, confrontés entre eux, explorés à l'aide des listes de spécificités puis interprétés à la lumière des connaissances contextuelles et des travaux socio-historiques disponibles. L'analyse qualitative, ici, succède à l'analyse quantitative et la conjonction des deux approches peut se révéler fructueuse (De Sousa, 2012b).
- 38 Dans le cas présent, seule sera mise à profit l'analyse arborée. Celle-ci permet de synthétiser sur un graphique les proximités et les distances<sup>22</sup> lexicales entre nos différents échantillons, et partant d'esquisser à grands traits une généalogie du discours révolutionnaire latino-américain à l'époque contemporaine.

- 39 L'analyse arborée (figure 6) distingue des polarités lexicales dans le discours révolutionnaire latino-américain : elle oppose principalement le discours guérillero (haut de l'arbre) et le discours indépendantiste (bas de l'arbre), avec, entre ces deux discours très marqués, quelques branches correspondant au discours communiste (Mella et Recabarren, respectivement fondateurs du Parti communiste cubain<sup>23</sup> et du Parti socialiste chilien), au discours singulier de l'EZLN – guérilla originale se revendiquant d'une multiplicité d'influences et parfois qualifiée de « post-marxiste » ou « post-moderne » – ou encore au discours de Guevara.
- 40 Sur l'arbre ci-après, on distingue nettement comment le M26 – mouvement de guérilla de Fidel Castro avant sa prise de pouvoir<sup>24</sup> –, puis Fidel Castro, s'intègrent à la branche du discours révolutionnaire indépendantiste et comment Fidel Castro, tout en y étant rattaché, en est le dernier bourgeon : après Fidel Castro, c'est une autre tradition discursive qui va s'implanter. En effet, si Fidel Castro – tout comme le M26, nous l'avons dit – reste fermement rattaché à la tradition discursive des Révolutions nationales et d'indépendance, c'est plutôt dans la tradition, plus récente, du discours communiste – déjà présent avant Fidel Castro et qui avait commencé à se développer en Amérique latine dans les années 20 et 30 (Mella, Racabarren) – qu'iront s'inscrire les révolutionnaires « post-castristes ». Le castrisme semble donc marquer un véritable tournant dans le discours révolutionnaire latino-américain des 19<sup>ème</sup> et 20<sup>ème</sup> siècles.
- 41 Bien entendu, l'analyse mériterait d'être prolongée : si l'on n'est guère surpris de trouver les uns aux côtés des autres les mouvements de guérillas fortement idéologisés (PCP-SL, MRTA, Santucho, EPR) et si, à l'opposé, la confluence entre les discours de Bolivar et Marti, de Zapata et Sandino confirment l'intuition du spécialiste, qu'en est-il de Salvador Allende, président socialiste élu, dont le discours s'apparente plus aux discours des guérillas des années 70 qu'au discours de Fidel Castro ou d'Hugo Chavez, tout deux chefs d'État ? Ou encore de Guevara qui semble se dégager de tout proche parentage ?

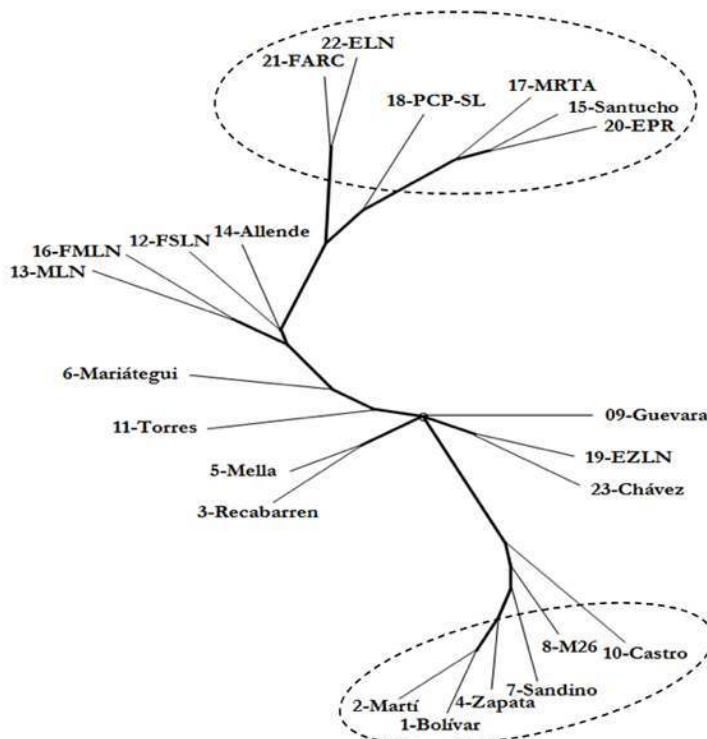


Figure 6 : Analyse arborée de la distance intertextuelle (calcul sur les formes graphiques)



## Conclusion

- 42 Au terme de ce bref parcours au sein du discours révolutionnaire latino-américain, nous espérons avoir montré la complémentarité entre statistique textuelle, théorie du discours et histoire sociale. Bien entendu, les méthodes quantitatives nécessitent un protocole et une réflexion préalable qui n'ont pas pu être exposés ici et il faut se garder d'une utilisation naïve de celles-ci. Les données doivent être construites et les faits linguistiques rendus mesurables. De plus, comme le remarquait Albert Soboul, « le fait étant mesurable, encore doit-il être en même temps qualifié et apprécié » (1967 : 17). Car le sens ne préexiste pas au corpus ; il n'émane du corpus que dans la mesure où ce dernier est amené à la vie par le chercheur. C'est pourquoi la constitution du corpus revêt un intérêt stratégique dans la mesure où toute modification de celui-ci est susceptible d'altérer les résultats statistiques sur lesquels s'appuiera l'interprétation. Pour l'historien, « c'est une révolution de la conscience historiographique » (Furet 1982 : 65).
- 43 Les chantiers ouverts par la complémentarité du quantitatif et du qualitatif en analyse du discours sont immenses : « pour la première fois dans l'histoire, la *totalité* des énoncés d'une société, appréhendée dans la multiplicité de leurs genres, est appelée à devenir objet d'étude » (Charaudeau & Maingueneau (dir.) 2002 : 45) et les méthodes quantitatives informatisées ne peuvent être laissées en marge de ce phénomène. Plus précisément, la complémentarité entre statistique textuelle, théorie du discours et histoire sociale doit mener à des études systématiques de l'aspect idéologique manifeste ou latent des discours des idéologies politiques, à travers les séries textuelles chronologiques adéquates, à en réviser les configurations et les périodisations, et à y déceler les sédimentations conceptuelles.

---

## BIBLIOGRAPHIE

- Amossy R. & Herschberg Pierrot A., 2009. *Stéréotypes et clichés : langue, discours, société*. Paris : Armand Colin.
- Angouri J., 2010. Quantitative, qualitative or both? Combining methods in linguistic research, in *Research methods in linguistics*, Litosseliti, L. (éd.). Londres : Continuum International Publishing Group. 29-45.
- Armony V., 1997. El análisis de datos cualitativos en ciencias sociales : nuevos enfoques y herramientas. *Revista de Investigaciones Folklóricas* 12. 9-16.
- Armony V., 2005. Aportes teórico-metodológicos para el estudio de la producción social de sentido a través del análisis del discurso presidencial. *Revista Argentina de Sociología* 3 (4). 32-54.
- Bergounioux A. et al., 1982. *La parole syndicale. Études du vocabulaire confédéral des Centrales ouvrière françaises, 1971-1976*. Paris : PUF.
- Bloch M., 1952. *Apologie pour l'histoire ou métier d'historien*. Paris : Armand Colin.

- Charaudeau P. & Maingueneau D. (dir.), 2002. *Dictionnaire d'analyse du discours*. Paris : Seuil.
- Chaunu P. & Chaunu H., 1962. Le climat des rapports franco-espagnols à Cadix dans la seconde moitié du XVII<sup>e</sup> siècle. Histoire sérielle et psychologie collective, problème de méthode. In *Mélanges Bataillon*. Bordeaux. 19-29.
- Chaunu P., 1955-1960. *Séville et l'Atlantique (1504-1650) (12 vol.)*. Paris : SEVPEN.
- Chaunu P., 1964. Histoire quantitative ou histoire sérielle. *Cahiers Vilfredo Pareto* 3. 165-176.
- Chaunu P., 1978. *Histoire quantitative, histoire sérielle*. Paris : Armand Colin/EHESS.
- Darbellay F., 2005. *Interdisciplinarité et transdisciplinarité en analyse des discours. Complexité des textes, intertextualité et transtextualité*. Genève : Slatkine.
- De Sousa S., 2010. Parentages et proximités segmentales dans le discours révolutionnaire latino-américain de Bolivar à la Révolution Bolivarienne (1810-2009). In S. Bolasco et al. (éd.), *Statistical analysis of textual data. Proceedings of 10th International Conference JADT 2010*. Milan : LED. 697-708.
- De Sousa S., 2012a. Peuple et populisme chez Hugo Chávez et Evo Morales. In M. Donot & M. Pordeus Ribeiro (éds), *Discours politiques en Amérique latine : imaginaires et représentations*. Paris : L'Harmattan. 69-84.
- De Sousa S., 2012b. À l'épreuve des temps... Temps lexical et temps politique dans le discours de Fidel Castro. In A. Dister et al. (éds), *JADT'12. Actes des 11èmes Journées internationales d'Analyse statistique des Données Textuelles*. Liège : LASLA-SESLA. 337-349.
- Détrie C., Siblot P. & Vérine B. (dir.), 2001. *Termes et concepts pour l'analyse du discours. Une approche praxématique*. Paris : Honoré Champion.
- Dörnyei Z., 2007. *Research methods in applied linguistics. Quantitative, qualitative, and mixed methodologies*. Oxford : Oxford University Press.
- Duchastel J. et Armony V., 1993. Un protocole de description des discours politiques. In *Secondes journées internationale d'analyse statistique de données textuelles*. Paris : Telecom. 159-183.
- Febvre L., 1992. *Combats pour l'histoire*. Paris : Armand Colin.
- Fiala P., 1994. L'interprétation en lexicométrie. Une approche quantitative des données lexicales. *Langue française* 103. 113-122.
- Foucault M., 1969. *L'archéologie du savoir*. Paris : Gallimard.
- Furet F., 1982. Le quantitatif en histoire. In *L'Atelier de l'histoire*. Paris : Flammarion.
- Geffroy A., Guilhaumou J. & Salem A., 1975. L'Histoire sur mesures ou pour une statistique du discours. *Bulletin du Centre d'analyse du discours de l'université de Lille* 3 2. Lille : PUL. 15-60.
- Guilbert T., 2007. *Le discours idéologique ou la force de l'évidence*. Paris : L'Harmattan.
- Guilhaumou J., 2002. Le corpus en analyse de discours : perspective historique. *Corpus* 1. 21-49.
- Guilhaumou J., 2006. *Discours et événement. L'histoire langagière des concepts*. Besançon : PUF-C.
- Guiraud P., 1960. *Problèmes et méthodes de la statistique linguistique*. Paris : PUF.
- Habert B., Nazarenko A. & Salem A., 1997. *La linguistique de corpus*. Paris : Armand Colin.
- Labbé D. & Monière D., 2003. *Le discours gouvernemental : Canada, Québec France (1945-2000)*. Paris : Honoré Champion.
- Lafon P., 1984. *Dépouillements et statistiques en lexicométrie*. Paris : Honoré Champion.
- Lebart L. et Salem, A., 1994. *Statistique textuelle*. Paris : Dunod.

- Luong X. (dir.), 2003. *Corpus 3. La distance intertextuelle*.
- Martinez W., 2003. *Contribution à une méthodologie de l'analyse des cooccurrences lexicales multiples dans les corpus textuels*. Thèse de doctorat en Sciences du Langage. Université de Paris 3, A. Salem (dir.).
- Mayaffre D., 2000. *Le poids des mots. Le discours de gauche et de droite dans l'entre-deux-guerres*. Maurice Thorez, Léon Blum, Pierre-Etienne Flandin et André Tardieu. Paris : Honoré Champion.
- Mayaffre D., 2002. L'Herméneutique numérique [version électronique]. Première publication : *L'Astrolabe. Recherche littéraire et Informatique*. Accessible à l'URL : <http://www.uottawa.ca/academic/arts/astrolabe/articles/art0031.htm/Hermeneutique.htm>
- Mayaffre D., 2005. De la lexicométrie à la logométrie [version électronique]. Première publication : *L'Astrolabe. Recherche littéraire et Informatique*. Accessible à l'URL : <http://www.uottawa.ca/academic/arts/astrolabe/articles/art0048/Logometrie.htm>
- Mayaffre D., 2008. Quand "travail", "famille", "patrie" co-occurrent dans le discours de Nicolas Sarkozy. Étude de cas et réflexion théorique sur la cooccurrence. In S. Heiden (éd.), *JADT 2008, 9èmes journées internationales d'analyse statistique des données textuelles*. Lyon : PUL. 811-822.
- Mazière F., 2005. *L'Analyse du discours*. Paris : PUF.
- Ortiz J., 1999. *Julio Antonio Mella l'ange rebelle. Aux origines du communisme cubain*. Paris : L'Harmattan.
- Pêcheux M., 1969. *Analyse automatique du discours*. Paris : Dunod.
- Prost A., 1969. Vocabulaire et typologie des familles politiques. *Cahiers de Lexicologie* 14. 115-126.
- Prost A., 1974. *Vocabulaire des proclamations électorales de 1881, 1885 et 1889*. Paris : PUF.
- Prost A., 1996. Les mots. In R. Rémond (dir.), *Pour une histoire politique*. Paris : Seuil. 255-285.
- Rastier F., 2011. *La mesure et le grain. Sémantique de corpus*. Paris : Honoré Champion.
- Robin R., 1973. *Histoire et linguistique*. Paris : Armand Colin.
- Salem A., 2006. Proximités segmentales. In J.-M. Viprey et al. (comp.), *JADT'06. Actes des 8èmes Journées internationales d'Analyse statistique des Données Textuelles*. Besançon : PUF-C. 843-853.
- Soboul A., 1967. Description et mesure en histoire sociale. In *L'Histoire sociale. Sources et méthodes, Colloque de l'ENS de Saint-Cloud (15-16 mai 1965)*. Paris : PUF.
- Tournier M., 1980. D'où viennent les fréquences de vocabulaire ? La lexicométrie et ses modèles. *Mots* 1. 189-209.
- Tournier M., 1987. Cooccurrences autour de travail (1971-1976). *Mots* 14. 89-123.
- Winock M., 1996. Les idées politiques. In R. Rémond (dir.), *Pour une histoire politique*. Paris : Seuil. 233-253.

## NOTES

1. Nous mettons entre guillemets cette première occurrence de discipline pour souligner le caractère problématique que peut prendre, à plus forte raison au sein d'un débat épistémologique, la division du champ du savoir en disciplines. Cf. par exemple Darbellay (2005).
2. Voir Febvre (1992 : 147-200) pour des articles sur ce sujet écrits entre 1906 et 1926. Ou encore Marc Bloch qui affirmait dans son *Apologie pour l'histoire ou métier d'historien* : « La liste des

"disciplines auxiliaires" dont nous proposons l'enseignement à nos débutants est beaucoup trop courte. Des hommes qui, la moitié du temps, ne pourront atteindre les objets de leurs études qu'à travers les mots, par quel absurde paralogisme leur permet on, entre autres lacunes, d'ignorer les acquisitions fondamentales de la linguistique ? » (1952 : 28).

3. Voir par exemple les travaux de Pierre Chaunu et les débats qu'ils ont pu susciter (Chaunu, 1978 ; 1955-1960, 12 vol.), mais aussi certains travaux de Fernand Braudel ou d'Emmanuel Leroy Ladurie.

4. L'histoire des prix, puis l'histoire du climat, par les séries linéaires que constituent les données qu'elles étudient, ont été pionnières dans le développement de l'histoire quantitative.

5. François Furet cité par Soboul (1967 : 16).

6. Nous nous limiterons, ici, à illustrer cette articulation plutôt qu'à la théoriser.

7. Le corpus a été dépouillé en formes graphiques.

8. Soit une fréquence relative de 14 pour 10000.

9. Par exemple, *pueblo* [4508 occ.], *lucha* [2948 occ.], *revolución* [2125 occ.] ou *masa* [1239 occ.], *república* [782 occ.], *socialismo* [718 occ.], *proletariado* [650 occ.], *trabajadores* [642 occ.].

10. Les nombres sur les branches indiquent la fréquence absolue du paradigme, des sous-paradigmes et des items dans le corpus. On ne détaillera pas ici la fréquence des sigles *EEUU* et *USA* qui ont été associés à la forme *Estados Unidos* dès le dépouillement du corpus. Notons que l'on trouve, dans certains textes, *norteamérica* sous la forme *Norte América* : celle-ci a été associée à *Norteamérica* dès le dépouillement. Par ailleurs, il est clair que les anaphores pronominales ne sont pas prises en compte ainsi que certaines lexies tel *Washington*, désignant parfois les États-Unis ; nous nous sommes contenté pour le moment de prendre en compte les désignants les plus fréquents et dont l'unicité référentielle ne peut être contestée.

11. Sur les cooccurrences : pour une réflexion méthodologique, cf. (Lafon, 1984) et (Martinez, 2003) ; pour des exemples d'application au discours politique, cf. (Tournier, 1987) et (Mayaffre, 2008).

12. On considérera ici le stéréotype comme « une représentation sociale, un schème collectif figé » (Amossy & Herschberg Pierrot 2009 : 64).

13. Il s'agit bien entendu d'un extrait regroupant les formes les plus spécifiques statistiquement (le haut de la liste), dont ont été exclu les verbes et les mots-outils (à l'exception de « contra »).

14. Dont les indices de probabilité sont compris entre  $10^{-50}$  et  $10^{-10}$ , c'est-à-dire que la probabilité que ces co-occurents soient associés si étroitement au paradigme des États-Unis dans ce corpus est très largement inférieure à 1 chance sur 10 milliards. Ce qui indique un lien très fort.

15. C'est la récurrence d'attributs qui, au sein d'un discours, semblent indispensables à la construction d'un thème ou d'une figure donnés qui amène à détecter le stéréotype au sens où nous l'entendons ici (cf. Amossy & Herschberg Pierrot 2009 : 106-110).

16. On entend par *doxa* « les évidences partagées, les stéréotypes, les croyances et les préjugés d'une communauté particulière ou d'un ensemble de communautés » (Guilbert 2007 : 122).

17. À la suite du *Dictionnaire d'analyse du discours*, on désigne par *formule* « une expression lexicale, le plus souvent un syntagme nominal ou une collocation à caractère néologique, qui renvoie à une notion ayant joué sur le plan idéologique un rôle fondateur et actif dans une situation historique » (Charaudeau & Maingueneau (dir.) 2002 : 274). Ce qui caractérise en particulier une formule, c'est son usage répété et sa circulation au sein des discours.

18. L'analyse du discours considère que « tout discours est traversé par l'interdiscursivité » c'est-à-dire qu'« il a pour propriété d'être en relation multiforme avec d'autres discours, d'entrer dans l'interdiscours » (Charaudeau & Maingueneau (dir.) 2002 : 324). La présence, au sein d'un discours, de formules circulant dans l'espace public peut être considérée comme un fait interdiscursif.

19. Notons, en comparaison, qu'une spécificité positive de 6, par exemple, signifie que, compte tenu des caractéristiques quantitatives du corpus et de ses parties ainsi que de la fréquence de

l'item observé, ce dernier avait une probabilité de  $10^{-6}$ , soit 1 chance sur 1 million, d'avoir une fréquence aussi élevée dans la partie du corpus considérée. Dans le cas présent, une spécificité positive de 28 signifie que ces segments avaient une probabilité de  $10^{-28}$ , soit une chance sur  $10^{28}$  d'avoir une fréquence aussi élevée dans les parties concernées (Staline, Trotski...) : en d'autres termes, de telles fréquences, hautement improbables, sont statistiquement remarquables.

20. « Ce "néologisme" du laboratoire de Saint-Cloud recouvre la question des éloignements ou rapprochements, voire des oppositions ou apparentements que l'on peut inférer, au plan formel, des emplois lexicaux globaux de nos locuteurs. » Il ne s'agit pas d'une « parenté de contenu » car « s'en tenir au "plan formel" signifie qu'on ne jugera d'un parentage que sur des critères de ressemblance graphique ou quantitative. » (Bergounioux et al., 1982 : 70).

21. Ou d'autre unité linguistique (catégories grammaticales...).

22. Le calcul de la distance (appelée parfois *connexion des textes* ou, lorsque qu'il s'agit de la *distance lexicale* et non plus de la *distance intertextuelle*, c'est-à-dire que seules les présences/ absences et non plus les fréquences des formes sont prises en compte, *connexion des vocabulaires*) vise à indiquer à quel point deux textes (ou ensembles de textes ou échantillons) partagent leur vocabulaire : quelle proportion de mots ou quelle surface textuelle leur est propre et quelle proportion de mots ou quelle surface textuelle leur est commune. Des algorithmes complexes procèdent à des calculs et aboutissent à des tableaux ou à des représentations graphiques. Le graphique 3 représente ainsi la distance entre chacun des échantillons pris deux à deux et le graphique 4, la distance entre l'échantillon M26 et chacun des 22 autres échantillons. Cf. (Luong (dir.), 2003).

23. Le premier Parti communiste cubain et non pas celui fondé par Fidel Castro en 1965. Cf. (Ortiz, 1999).

24. Tous les textes du M26 inclus ici peuvent être considérés *a priori* comme ayant été écrit de la main de Fidel Castro.

## RÉSUMÉS

L'articulation du qualitatif et du quantitatif, sa pertinence et les modalités de sa mise en œuvre ont longtemps été questionnées hors de l'analyse du discours. Il nous apparaît que, dans le cadre particulier de l'analyse du discours, quantitatif et qualitatif ont intérêt à être mobilisés conjointement en articulant théorie du discours et méthodes statistiques. C'est cette position qui sera défendue à travers une réflexion épistémologique illustrée ensuite d'exemples d'analyses tirées de notre parcours de recherche.

The articulation between qualitative and quantitative approaches, its relevance and the modalities of its implementation have long been questioned outside of discourse analysis. It appears that, in discourse analysis in particular, qualitative and quantitative could be well advise to be called up jointly by articulating discourse theory and statistical methods. This position will be defended through an epistemological reflection, then illustrated by examples drawn from our framework.

## INDEX

**Keywords** : Discourse Analysis, Latin American revolutionary discourse, quantitative, qualitative, Textual Statistic

**Mots-clés** : Analyse du discours, discours révolutionnaire latino-américain, quantitatif, qualitatif, statistique textuelle

## AUTEUR

**SERGE DE SOUSA**

CRIT – UFR SLHS – BesançonSYLED-CLA<sup>2</sup>T – Université Paris III